JOURNAL OF MULTIVARIATE ANALYSIS 7, 235–248 (1977)

A Multidimensional Newton–Raphson Method and Its Applications to the Existence of Asymptotic F_n-Estimators and Their Stochastic Expansions

R. MICHEL

Mathematisches Institut der Universität Köln, Federal Republic of Germany

Communicated by L. Schmetterer

Given a suitable function F_n we define a class of estimators called asymptotic F_n-estimators (i.e., estimators which approximate the solution of $F_n(\theta) = 0$). It is proved that this class is nonvoid if appropriate regularity conditions are fulfilled and if one has at hand a suitable initial estimator. Furthermore, it is shown that F_n-estimators admit a stochastic expansion (which enables to give results on asymptotic expansions for the distribution of these estimators).

1. INTRODUCTION AND NOTATIONS

Up to now the main tools in obtaining results on asymptotic expansions for the distribution of maximum likelihood estimators have been a stochastic expansion of the estimator, i.e., a polynomial in the standardized derivatives of $\theta \to \sum_{i=1}^n \log p(x_i, \theta)$ which is sufficiently close to the maximum likelihood estimator, in connection with results giving asymptotic expansions for the distribution of such polynomials. For one-dimensional parameters this was first performed by Linnik and Mitrofanova [5]. A first attempt for the case of vector parameters has been made by Mitrofanova [7] under very restrictive regularity conditions. A rigorous proof of a more general result making clear, too, the two main steps of the approach has been given by Chibisov [1, 2]. For a result concerning asymptotic maximum likelihood estimators (which include, under appropriate regularity conditions, maximum likelihood estimators), see also the paper by Michel [6]. In all of these papers, the structure of the stochastic expansion of the estimators is not very transparent (e.g., in [2, 7], it is obtained from a formal series-expansion of the solution of the likelihood equation).

In this paper we show that a suitable Newton–Raphson iteration procedure

Received September 24, 1976.

AMS 1970 subject classifications: 62E20, 62H99.

Key words and phrases. Iteration procedure, stochastic expansions, maximum likelihood estimators, expansion for the distribution.

ISSN 0047-259X

yields, in a general case, the existence of suitable estimators together with their stochastic expansions. Previous results are improved in several respects:

(i) A class of estimators whose distribution functions admit (under suitable regularity conditions) an asymptotic expansion is presented. This class includes minimum contrast estimators and asymptotic minimum contrast estimators (hence, it includes especially (asymptotic) maximum likelihood estimators). Furthermore, our approach is not limited to the i.i.d. case.

(ii) Complete and new information is given about the structure of the functions constituting the stochastic approximation of the estimators. The approximation can be obtained from a recurrence formula in connection with a portion of a geometric series (see Remark 1 in Section 2). In the one-dimensional case, the stochastic expansion is presented in a closed form (see Remark 2 in Section 2). Furthermore, it appears that both the estimator and its stochastic expansion may be obtained by an application of one general concept. Both are approximations of the solution of $F_n(\theta) = 0$, where F_n is the function being considered (e.g., F_n may be equal to $\theta \rightarrow n^{-1} \sum_{i=1}^{n} (\partial/\partial\theta) \log p(x_i, \theta)$ if we deal with asymptotic maximum likelihood estimators).

If we take, e.g., the function $f_{(2)}^{(r)}$ of Remark 1, which fulfills properties (5) and (6) of Lemma 2, and if we know a suitable "initial estimator" $T_n^{(0)}$ (see Definition 1 in Section 3) then

$$T_n(x) = f_{(2)}^{(r)}(T_n^{(0)}(x))$$

is an estimator of our class (an asymptotic F_n-estimator) and

$$S_n^{(r)}(x, \theta) = f_{(2)}^{(r)}(\theta)$$

is its stochastic expansion. (Recall that $f_{(2)}^{(r)}$ depends on x and n through F_n.)

(iii) As an extension of the known results, it is shown that

$$\sup\{\| T_n(x) - S_n^{(r)}(x, \tau)\| : \| \tau - \theta \| < \epsilon_K n^{-1/2}(\log n)^2\}$$

is small with high probability, when θ is the true parameter value. This result may be of special interest in testing hypotheses.

The improvement procedure based on the Newton–Raphson method with quadratic convergence (see Lemma 5 in Section 2) applied to the numerical solution of the maximum likelihood equation was proposed as early as 1925 by Fisher [3] and was used by LeCam [4] to obtain asymptotically normal estimators of minimal variance. It has been taken up again in the papers of Pfanzagl [9] and Michel [6].

Our improvement procedure (see Lemma 4 of Section 2) is a Newton–Raphson iteration procedure with a higher degree of convergence (and is in fact "equivalent" to a sufficiently high number of steps of the former; see Lemma 5 of Section 2).

The following notations are used in this paper:

$\| \ \|$ denotes Euclidean norm, i.e., for any multilinear form H^j on $\mathbb{R}^k$,

$$H^j = (h_{i_1,\dots,i_j})_{i_1,\dots,i_j=1,\dots,k},$$

$$\| H^j \| = \left(\sum_{i_1,\dots,i_j=1}^{k} h^2_{i_1,\dots,i_j} \right)^{1/2}.$$

Observe that $j = 1$ means vectors and $j = 2$ denotes matrices.

Furthermore, given a function $F: \mathbb{R}^k \to \mathbb{R}^k$, we write $D^jF(x)$ and $DF(x)$ (in the case $j = 1$) for the multilinear form with entries

$$D_{i_1,\dots,i_j}F(x), \qquad i_1,\dots, i_j = 1,\dots, k$$

where

$$D_{i_1,\dots,i_j}F(x) = \frac{\partial^j}{\partial x_{i_1} \cdots \partial x_{i_j}} F(x).$$

Finally, for $x_0 \in \mathbb{R}^k$ and $\delta > 0$ let

$$K_\delta(x_0) = \{x \in \mathbb{R}^k: \| x - x_0 \| \leqslant \delta\}.$$

2. An Appropriate Newton–Raphson Method

The results of this section are given in some lemmas dealing with the existence and the properties of an appropriate iteration procedure. Lemmas 1 and 3 are more or less technical. Lemma 2 states the properties of two functions $f_{(1)}^{(r)}$ and $f_{(2)}^{(r)}$ which are in a certain sense equivalent to a given iteration procedure function $f^{(r)}$. In the statistical applications given in Section 3, $f_{(1)}^{(r)}$ is reserved for the definition of the estimator and $f_{(2)}^{(r)}$ is used for defining its stochastic expansion. (Observe that we can take $f_{(2)}^{(r)} = f_{(1)}^{(r)}$.) In Lemma 4 we show that (under appropriate regularity conditions) there exists a function $f^{(r)}$ with the properties assumed in Lemma 2. In the first two remarks following Lemma 4 a well-suited function $f_{(2)}^{(r)}$ is defined. (The second remark deals with the one-dimensional case and here we give a closed form of the function $f_{(2)}^{(r)}$.) In Lemma 5 we finally show that the usual Newton–Raphson procedure (used, e.g., in [6, 9] for the definition of asymptotic maximum likelihood estimators) gives rise to a function $f_{(1)}^{(r)}$.

The reader should not be scared by the many constants in these lemmas and the conditions on the constants. One should always bear in mind that in the satistical applications of Section 3 these constants (i.e., α, δ, ϵ, γ, η, τ, ω) are of the form

$$n^{-1/2}(\log n)^{\gamma_K} \text{ respectively } \epsilon_K n^{-1/2}(\log n)^2$$

for suitable constants ϵ_K, $\gamma_K > 0$, and all the conditions imposed reduce to only three conditions, namely, that $n \in \mathbb{N}$ respectively $\gamma_K > 0$ should be sufficiently large and $\epsilon_K > 0$ sufficiently small.

In the following we assume that $F\colon \mathbb{R}^k \to \mathbb{R}^k$ is a differentiable function.

LEMMA 1. *Assume that there exists a regular matrix Γ_0 and a constant $\delta > 0$ such that*

$$\sup_{x \in K_\delta(x_0)} \| \Gamma_0^{-1} \| \, \| DF(x) - \Gamma_0 \| < \tfrac{1}{2}$$

and

$$\| \Gamma_0^{-1} \| \, \| F(x_0) \| < \tfrac{1}{2}\delta.$$

Then there exists $x_* \in K_\delta(x_0)$ with $F(x_*) = 0$.

Proof. Let $H(x) = x - \Gamma_0^{-1}F(x)$. We have

$$\sup_{z \in K_\delta(x_0)} \| DH(z) \| \leqslant \sup_{z \in K_\delta(x_0)} \| \Gamma_0^{-1} \| \, \| DF(z) - \Gamma_0 \| < \tfrac{1}{2} \tag{1}$$

and therefore $x \in K_\delta(x_0)$ implies

$$\begin{aligned} \| H(x) - x_0 \| &\leqslant \| \Gamma_0^{-1} \| \, \| F(x_0) \| + \| x - \Gamma_0^{-1}F(x) - (x_0 - \Gamma_0^{-1}F(x_0)) \| \\ &\leqslant \tfrac{1}{2}\delta + \| x - x_0 \| \sup_{z \in K_\delta(x_0)} \| DH(z) \| \leqslant \delta. \end{aligned} \tag{2}$$

By (1) and (2), the fixed-point theorem for contracting maps implies the existence of $x_* \in K_\delta(x_0)$ with $H(x_*) = x_*$, i.e., $F(x_*) = 0$.

LEMMA 2. *Let the assumptions of Lemma* 1 *be fulfilled. Assume that there exists $r \in \mathbb{N}$, a differentiable function $f^{(r)}\colon K_\delta(x_0) \to \mathbb{R}^k$, and constants ϵ, $\tau > 0$ with $\epsilon < 1/6$ and $\tau^2 < \delta/6$ such that*

$$F(x) = 0 \qquad \textit{implies} \qquad f^{(r)}(x) = x,$$

$$\sup_{x \in K_\delta(x_0)} \| f^{(r)}(x) - H(x) \| < \tau^2, \tag{3}$$

where $H(x) = x - \Gamma_0^{-1}F(x)$, and

$$\sup_{x \in K_\delta(x_0)} \| Df^{(r)}(x) \| < \epsilon^r.$$

Let $f_{(i)}^{(r)}\colon K_\delta(x_0) \to \mathbb{R}^k$, $i = 1, 2$, be functions and $\eta \in [0, 1)$ a constant with $\eta^2 < \delta/6$ such that

$$\sup_{x \in K_\delta(x_0)} \| f_{(i)}^{(r)}(x) - f^{(r)}(x) \| \leqslant \eta^{r+1}. \tag{4}$$

Then,

$$\sup_{x,z\in K_\delta(x_0)} \| f_{(1)}^{(r)}(x) - f_{(2)}^{(r)}(z)\| < 2(\eta^{r+1} + \delta\epsilon^r) \tag{5}$$

and

$$\sup_{x\in K_\delta(x_0)} \| F(f_{(1)}^{(r)}(x))\| < 3 \,\| \Gamma_0 \|(\eta^{r+1} + \delta\epsilon^r). \tag{6}$$

Proof. (i) We have for $x, z \in K_\delta(x_0)$,

$$\begin{aligned}\| f_{(1)}^{(r)}(x) - f_{(2)}^{(r)}(z)\| &\leqslant \| f_{(1)}^{(r)}(x) - f^{(r)}(x)\| + \| f^{(r)}(x) - f^{(r)}(z)\| \\ &\quad + \| f^{(r)}(z) - f_{(2)}^{(r)}(z)\| \\ &\leqslant 2\eta^{r+1} + \| x - z\| \sup_{y\in K_\delta(x_0)} \| Df^{(r)}(y)\| \\ &< 2(\eta^{r+1} + \delta\epsilon^r),\end{aligned}$$

and this bound is independent of x and z.

(ii) At first we show that $x\in K_\delta(x_0)$ implies $f_{(1)}^{(r)}(x)\in K_\delta(x_0)$ (then this especially holds true for $f^{(r)}$):

We have for $x\in K_\delta(x_0)$,

$$\begin{aligned}\| f_{(1)}^{(r)}(x) - x_0 \| &\leqslant \| f_{(1)}^{(r)}(x) - f^{(r)}(x)\| + \| f^{(r)}(x) - f^{(r)}(x_0)\| \\ &\quad + \| f^{(r)}(x_0) - H(x_0)\| + \| H(x_0) - x_0 \| \\ &< \eta^{r+1} + \| x - x_0 \| \sup_{z\in K_\delta(x_0)} \| Df^{(r)}(z)\| + \tau^2 \\ &\quad + \| \Gamma_0^{-1} \| \, \| F(x_0)\| \\ &< \eta^2 + \delta\epsilon + \tau^2 + \delta/2 < \delta.\end{aligned}$$

Let $x_* \in K_\delta(x_0)$ be given with $F(x_*) = 0$ (see Lemma 1). Since $F(x_*) = 0$ implies $f^{(r)}(x_*) = x_*$, i.e., $F(f^{(r)}(x_*)) = 0$, we obtain for $x\in K_\delta(x_0)$,

$$\begin{aligned}\| F(f_{(1)}^{(r)}(x))\| &\leqslant \| F(f_{(1)}^{(r)}(x)) - F(f^{(r)}(x))\| \\ &\quad + \| F(f^{(r)}(x)) - F(f^{(r)}(x_*))\|.\end{aligned}$$

This together with

$$\sup_{y\in K_\delta(x_0)} \| DF(y)\| < \| \Gamma_0 \| + \tfrac{1}{2} \| \Gamma_0^{-1} \|^{-1} \leqslant (1 + \tfrac{1}{2}k^{-1/2}) \| \Gamma_0 \| \leqslant \tfrac{3}{2} \| \Gamma_0 \|$$

implies

$$\begin{aligned}\sup_{x\in K_\delta(x_0)} \| F(f_{(1)}^{(r)}(x))\| &< \tfrac{3}{2} \| \Gamma_0 \| \,(\eta^{r+1} + 2\delta\epsilon^r) \\ &\leqslant 3 \| \Gamma_0 \| \,(\eta^{r+1} + \delta\epsilon^r).\end{aligned}$$

LEMMA 3. *Assume that F is $(r+1)$-times differentiable and that there exists a regular matrix Γ_0 and constants α, $A_r > 0$ with $\alpha < \frac{1}{2}$ such that*

$$\| \Gamma_0^{-1} \| \, \| DF(x_0) - \Gamma_0 \| < \tfrac{1}{2}\alpha$$

$$\| D^j F(x_0) \| < A_r , \qquad j = 2, \ldots, r$$

$$\sup_{x \in K_\alpha(x_0)} \| D^{r+1} F(x) \| < A_r .$$

Then there exists a constant β_r, $B_r > 0$ (depending only on r and A_r) such that

$$\sup_{x \in K_\delta(x_0)} \| \Gamma_0^{-1} \| \, \| DF(x) - \Gamma_0 \| < \alpha \tag{7}$$

$$\sup_{x \in K_\delta(x_0)} \| D^j F(x) \| < B_r , \qquad j = 2, \ldots, r, \tag{8}$$

where

$$\delta = \alpha \min(\| \Gamma_0^{-1} \|^{-1} \beta_r , \tfrac{1}{2} \| \Gamma_0 \|^{-1}, 1). \tag{9}$$

Proof. Obvious.

LEMMA 4. *Let the conditions of Lemma 3 be fulfilled, and assume in addition that*

$$\| \Gamma_0^{-1} \| \, \| F(x_0) \| < \delta/2 \tag{10}$$

where δ is given by (9). Then there exists a function $f^{(r)}: K_\delta(x_0) \to \mathbb{R}^k$ fulfilling the conditions in (3) of Lemma 2 for all τ, $\epsilon > 0$ with

$$\left.\begin{aligned} &\tau \geqslant \alpha(2 + C_r)(1 + 2 \| \Gamma_0^{-1} \|)^{m_r} \qquad \textit{and} \qquad \tau^2 < \delta/6 \\ &\epsilon \geqslant \alpha[\Gamma_0 C_r (1 + 2 \| \Gamma_0^{-1} \|)^{m_r}]^{1/r} \qquad \textit{and} \qquad \epsilon < 1/6, \end{aligned}\right\} \tag{11}$$

where C_r and m_r are given in (21) below.

Proof. (i) F is injective on $K_\delta(x_0)$ and the inverse function $F^{-1}: F(K_\delta(x_0)) \to K_\delta(x_0)$ is $(r+1)$-times differentiable. To prove the first assertion set $H(x) = x - \Gamma_0^{-1} F(x)$ and let $x, y \in K_\delta(x_0)$ be given with $F(x) = F(y)$. Then by (7),

$$\| x - y \| = \| H(x) - H(y) \| \leqslant \| x - y \| \sup_{z \in K_\delta(x_0)} \| DH(z) \|$$

$$\leqslant \alpha \| x - y \| \leqslant \tfrac{1}{2} \| x - y \|,$$

which implies $x = y$.

The second assertion now follows from (7) and the assumptions. We furthermore have,

$$\sup_{x\in K_\delta(x_0)} \| DF(x)\| < \tfrac{3}{2} \| \Gamma_0 \|, \qquad \sup_{y\in F(K_\delta(x_0))} \| DF^{-1}(y)\| < 2 \| \Gamma_0^{-1} \|$$

and (12)

$$\sup_{y\in F(K_\delta(x_0))} \| \Gamma_0 DF^{-1}(y) - I \| < 2\alpha.$$

This immediately follows from (7) together with $\alpha < \frac{1}{2}$ and the Neumann series

$$DF^{-1}(y) = \Gamma_0^{-1} \sum_{m=0}^{\infty} [(\Gamma_0 - DF(x)\,|_{x=F^{-1}(y)})\, \Gamma_0^{-1}]^m. \tag{13}$$

(ii) For $j \in \mathbb{N}$ let $S^j(x) = (S^j_{ii_1\cdots i_j}(x))_{i,i_1,\ldots,i_j=1,\ldots,k}$, where

$$S^j_{ii_1\cdots i_j}(x) = D_{i_1\cdots i_j} F_i^{-1}(y)\,|_{y=F(x)}\,. \tag{14}$$

Notice that for $h \in \mathbb{R}^k$,

$$S^j(x)\, h^j = \left(\sum_{i_1,\ldots,i_j=1}^{k} S^j_{ii_1\cdots i_j}(x) \prod_{\nu=1}^{j} h_{i_\nu} \right)_{i=1,\ldots,k}. \tag{15}$$

For $h \in \mathbb{R}^k$ and a $k \times k$ matrix $M = (M_{ip})_{i,p=1,\ldots,k}$ define, furthermore,

$$S^{j+1}(x)\, h^j M = \left(\sum_{i_1,\ldots,i_{j+1}=1}^{k} S^{j+1}_{ii_1\cdots i_{j+1}}(x)\, M_{i_{j+1}p} \prod_{\nu=1}^{j} h_{i_\nu} \right)_{i,p=1,\ldots,p}. \tag{16}$$

Observe that (14)–(16) yield the following recurrence formula for the S^j:

$$S^1(x) = DF(x)^{-1}$$

and for $j \geqslant 1$, (17)

$$D[S^j(x)h^j] = S^{j+1}(x)\, h^j DF(x), \qquad h \in \mathbb{R}^k.$$

With the preceding notations we obviously have

$$D[S^j(x)\, F(x)^j] = S^{j+1}(x)\, F(x)^j\, DF(x) + jS^j F(x)^{j-1}\, DF(x). \tag{18}$$

(iii) After these preliminaries, we now define the function $f^{(r)}: K_\delta(x_0) \to \mathbb{R}^k$. Let

$$f^{(r)}(x) = x + \sum_{j=1}^{r} \frac{(-1)^j}{j!}\, S^j(x)\, F(x)^j, \qquad x \in K_\delta(x_0). \tag{19}$$

Notice that the components $f_i^{(r)}$ of $f^{(r)}$ are nothing other than a portion of the Taylor series of $y \to F_i^{-1}(y)$ around $F(x)$, taken for the particular value $y = 0$.

From the fact that the elements constituting $S^j(x)$ are polynomials in $DF(x)^{-1}$ and the derivatives of $F(x)$ up to the order j we immediately obtain that $f^{(r)}$ is differentiable with (see (18) and (19))

$$Df^{(r)}(x) = \frac{(-1)^r}{r!} S^{r+1}(x)\, F(x)^r\, DF(x). \tag{20}$$

(iv) Using (12) and the arguments preceding (20), we obtain the existence of a constant $C_r > 0$ (depending only on r and A_r) and a natural number m_r such that for every $h \in \mathbb{R}^k$, every $(k \times k)$ matrix M and all $j = 1,\ldots, r$

$$\sup_{x \in K_\delta(x_0)} \| S^j(x)\, h^j \| \leqslant C_r(1 + 2 \| \Gamma_0^{-1} \|)^{m_r} \| h \|^j \tag{21}$$

and

$$\sup_{x \in K_\delta(x_0)} \| S^{j+1}(x)\, h^j M \| \leqslant C_r(1 + 2 \| \Gamma_0^{-1} \|)^{m_r} \| h \|^j \| M \|. \tag{22}$$

Furthermore, (7) and (10) together with $\| \Gamma_0^{-1} \|^{-1} \leqslant \| \Gamma_0 \|$ imply

$$\sup_{x \in K_\delta(x_0)} \| F(x) \| < 2 \| \Gamma_0 \|\, \delta \leqslant \alpha. \tag{23}$$

The conditions on ϵ, τ now follow immediately from (9), (17), and (19)–(23).

Remark 1. The function $f^{(r)}(x)$ is not well suited for our purposes, as the inverse of the matrix $DF(x)$ occurs in the representation of $f^{(r)}(x)$. A candidate for $f^{(r)}_{(2)}(x)$ [resp. $f^{(r)}_{(1)}(x)$] is obtained by the following device:

Replace $DF(x)^{-1}$ in $f^{(r)}$ by portions of the series

$$DF(x)^{-1} = \Gamma_0^{-1} \sum_{m=0}^{\infty} [(\Gamma_0 - DF(x))\, \Gamma_0^{-1}]^m, \tag{24}$$

(whose mth term is bounded on $K_\delta(x_0)$ by $\| \Gamma_0^{-1} \|\, \alpha^m$ (see (7))) in such a way that the remainder terms are bounded by

$$E_r(1 + \| \Gamma_0^{-1} \|)^{q_r}\, \alpha^{r+1} \tag{25}$$

for suitable $q_r \in \mathbb{N}$ and a constant E_r (both depending only on r and A_r). Observe that $\sup_{x \in K_\delta(x_0)} \| F(x) \| < \alpha$ by (23). Then (4) is fulfilled for every $\eta > 0$ such that $\eta \geqslant \alpha[E_r(1 + \| \Gamma_0^{-1} \|)^{q_r}]^{1/(r+1)}$ and $\eta^2 < \delta/6$

We illustrate this construction for the case $r = 3$. To simplify our notations we use the following convention: If in a product an index occurs twice, then this means the summation over this index from 1 to k, e.g.,

$$a_{ij} b_j \quad \text{means} \quad \sum_{j=1}^{k} a_{ij} b_j\,.$$

Let $(G_{pq})_{p,q=1,\dots,k}$ denote the inverse of $DF(x)$. Hence

$$G_{pq}D_rF_q(x) = \delta_{pr}\,.$$

Then we have

$$D_\nu G_{pq} = -G_{\alpha q}G_{p\beta}D_{\alpha\nu}F_\beta\,. \tag{26}$$

Let, furthermore,

$$S^jh^j = (S^j_{ii_1\cdots i_j}h_{i_1}\cdots h_{i_j})_{i=1,\dots,k}\,.$$

Then (17) yields

$$S^1_{ii_1} = G_{ii_1} \tag{27}$$

and for $j \geqslant 1$,

$$S^{j+1}_{ii_1\cdots i_{j+1}} = G_{\alpha i_{j+1}}D_\alpha S^j_{ii_1\cdots i_j}\,. \tag{28}$$

Hence,

$$\begin{aligned} S^1_{ii_1} &= G_{ii_1} \\ S^2_{ii_1i_2} &= -G_{i\alpha}G_{\beta i_1}G_{\gamma i_2}D_{\beta\gamma}F_\alpha \\ S^3_{ii_1i_2i_3} &= G_{i\alpha}G_{\beta i_1}G_{\gamma i_2}G_{\delta i_3}(3G_{\epsilon\nu}D_{\beta\delta}F_\nu D_{\epsilon\gamma}F_\alpha - D_{\beta\gamma\delta}F_\alpha). \end{aligned} \tag{29}$$

Let

$$(\Gamma_{ij})_{i,j=1,\dots,k} = \Gamma_0 \qquad \text{and} \qquad (\Lambda_{ij})_{i,j=1,\dots,k} = \Gamma_0^{-1}. \tag{30}$$

From (24) we have to take at most the portion

$$\Gamma_0^{-1} + \Gamma_0^{-1}(\Gamma_0 - DF(x))\,\Gamma_0^{-1} + \Gamma_0^{-1}[(\Gamma_0 - DF(x))\,\Gamma_0^{-1}]^2$$

whose entries are

$$\Lambda_{ij} + \Lambda_{i\alpha}(\Gamma_{\alpha\beta} - D_\beta F_\alpha)\,\Lambda_{\beta j} + \Lambda_{i\alpha}(\Gamma_{\alpha\beta} - D_\beta F_\alpha)\,\Lambda_{\beta\gamma}(\Gamma_{\gamma\delta} - D_\delta F_\gamma)\Lambda_{\gamma j}\,. \tag{31}$$

We finally set

$$u_i = \Lambda_{i\alpha}F_\alpha\,, \qquad u_{ij} = \Lambda_{i\alpha}(D_jF_\alpha - \Gamma_{\alpha j}). \tag{32}$$

Then we obtain from (19) and (29)–(32) the following result on the ith component of $f^{(3)}_{(2)}$:

$$\begin{aligned} f^{(3)}_{(2)i} = &-u_i + u_{i\beta}u_\beta - \tfrac{1}{2}\Lambda_{i\alpha}u_\beta u_\gamma D_{\beta\gamma}F_\alpha \\ &+ [\tfrac{1}{6}\Lambda_{i\alpha}D_{\beta\gamma\delta}F_\alpha - \tfrac{1}{2}\Lambda_{i\alpha}\Lambda_{\epsilon\nu}(D_{\epsilon\gamma}F_\alpha)\,D_{\beta\delta}F_\nu]\,u_\beta u_\gamma u_\delta \\ &- u_{i\beta}u_{\beta\delta}u_\delta + \Lambda_{i\alpha}u_{\beta\epsilon}u_\gamma u_\epsilon D_{\beta\gamma}F_\alpha \\ &+ \tfrac{1}{2}u_{i\epsilon}u_\beta u_\gamma\Lambda_{\epsilon\alpha}D_{\beta\gamma}F_\alpha\,. \end{aligned}$$

Remark 2. By some obvious but tedious calculations we derive in the case $k = 1$ the following closed form of the function $f_{(2)}^{(r)}$ constructed in Remark 1:

$$f_{(2)}^{(r)}(x) = x + \sum_{j=1}^{r} \sum_{i=1}^{j} \frac{1}{i!} \binom{j+i-2}{2i-2}$$

$$\times \left\{ \sum{}^{*} (-1)^{j_1+1} (2i-2-j_1)! \, [\Gamma_0^{-1} F^{(1)}(x)]^{j_1} \prod_{\nu=2}^{i} \frac{1}{j_\nu !} \left[\frac{1}{\nu !} F^{(\nu)}(x) \right]^{j_\nu} \right\}$$

$$\times [\Gamma_0^{-1} F(x)]^{i} \, [\Gamma_0^{-1}(\Gamma_0 - F^{(1)}(x))]^{j-i},$$

where $F^{(\nu)}(x)$ denotes the derivative $(d^\nu/dx^\nu)F(x)$ and where $\sum^*$ means summation over all $(j_1, \ldots, j_i) \in \{0, 1, \ldots, i-1\}^i$ with $\sum_{\alpha=1}^{i} j_\alpha = i - 1$ and $\sum_{\alpha=1}^{i} \alpha j_\alpha = 2i - 2$.

Remark 3. If $x_r \in \mathbb{R}^k$ is given such that x_r is sufficiently close to x_0 and such that

$$\| F(x_r) \| < \omega^{r+1} \tag{33}$$

for some $\omega > 0$ (i.e., if x_r approximates the solution of $F(x) = 0$), then x_r is in a certain sense equivalent of $f^{(r)}(x)$. To be precise, assume that the conditions of Lemmas 2 and 4 and (33) are fulfilled and that

$$x_r \in K_\delta(x_0),$$

where δ is given by (9). Then,

$$\sup_{x \in K_\delta(x_0)} \| x_r - f^{(r)}(x) \| < 2 \, \| \Gamma_0^{-1} \| \, (1 + 3 \, \| \Gamma_0 \|)[\max(\omega, \delta, \epsilon)]^{r+1}, \tag{34}$$

where $f^{(r)}$ is any function fulfilling (3) of Lemma 2. The proof in (34) is intuitively clear: The function F is locally Lipschitzian in both directions. We indicate the proof:

By the first lines of the proof of Lemma 4, F is injective on $K_\delta(x_0)$ and $F^{-1}: F(K_\delta(x_0)) \to K_\delta(x_0)$ is differentiable with (see (12))

$$\sup_{y \in F(K_\delta(x))} \| DF^{-1}(y) \| < 2 \, \| \Gamma_0^{-1} \|.$$

Hence, we have

$$\| x_r - f^{(r)}(x) \| = \| F^{-1}(F(x_r)) - F^{-1}(F(f^{(r)}(x))) \|$$

$$< 2 \, \| \Gamma_0^{-1} \| \, (\| F(x_r) \| + \| F(f^{(r)}(x)) \|),$$

which implies the assertion.

LEMMA 5. *Let the conditions of Lemma 2 be fulfilled and assume that (instead of the first condition of Lemma 1),*

$$\sup_{x\in K_\delta(x_0)} \| \Gamma_0^{-1}\| \, \| DF(x) - \Gamma_0 \| < \alpha, \qquad \textit{where} \quad \alpha < \tfrac{1}{2}.$$

Let $H^{(1)}(x) = H(x)$ $[= x - \Gamma_0^{-1}F(x)]$ *and for* $j \geqslant 1$, $H^{(j+1)}(x) = H(H^{(j)}(x))$, $x \in K_\delta(x_0)$. *Then* $f_{(1)}^{(r)}(x) = H^{(r)}(x)$ *fulfills* (4) *for every*

$$\eta \geqslant (1 + 6 \| \Gamma_0^{-1}\| \, \| \Gamma_0 \|)^{1/(r+1)} \max(\alpha, \delta, \epsilon, \tau) \textit{ with } \eta^2 < \delta/6.$$

Proof. We shall prove that for $j \in \mathbb{N}$,

$$\sup_{x\in K_\delta(x_0)} \| H^{(j)}(x) - f^{(r)}(x)\| < \alpha^{j-1}\tau^2 + 3 \| \Gamma_0^{-1}\| \, \| \Gamma_0 \| \, \delta\epsilon^r \sum_{\nu=0}^{j-2} \alpha^\nu. \tag{35}$$

Since $\alpha < \frac{1}{2}$, this then implies the assertion (by setting $j = r$.)
By assumption, we have for $x \in K_\delta(x)$ (case $j = 1$)

$$\| H^{(1)}(x) - f^{(r)}(x)\| = \| H(x) - f^{(r)}(x)\| < \tau^2.$$

As shown in the proof of Lemma 1, $H(K_\delta(x)) \subset K_\delta(x)$, i.e., $x \in K_\delta(x_0)$ implies $H^{(j)}(x) \in K_\delta(x)$ for every $j \in \mathbb{N}$. Therefore, we have for $j \geqslant 2$ and $x \in K_\delta(x_0)$,

$$\begin{aligned} \| H^{(j)}(x) - f^{(r)}(x)\| &\leqslant \| H(H^{(j-1)}(x)) - H(f^{(r)}(x))\| \\ &\quad + \| \Gamma_0^{-1} \| \, \| F(f^{(r)}(x))\| \\ &< \alpha \| H^{(j-1)}(x) - f^{(r)}(x))\| + 3 \| \Gamma_0^{-1} \| \, \| \Gamma_0 \| \, \delta\epsilon^r \end{aligned}$$

as $\sup_{z\in K_\delta(x_0)} \| DH(z)\| < \alpha$. (Observe that $x \in K_\delta(x_0)$ implies $f^{(r)}(x) \in K_\delta(x_0)$ by the first lines of the proof of Lemma 2(ii).) Furthermore, we have used Lemma 2(ii) with $\eta = 0$ and $f_{(1)}^{(r)} = f^{(r)}$. Hence, (35) for $j - 1$ implies (35) for j.

3. STATISTICAL APPLICATIONS

Let $(X, \mathscr{A})$ be a measurable space and $P_\theta \mid \mathscr{A}$, $\theta \in \Theta$, $\Theta \subset \mathbb{R}^k$ open, a family of probability measures.

DEFINITION 1. A measurable map $T_n^{(0)}: X \to \mathbb{R}^k$ is an initial estimator of order $s \in \mathbb{N}$ if for every compact subset K of Θ and every constant $c > 0$,

$$\sup_{\theta\in K} P_\theta\{x \in X : \| T_n^{(0)}(x) - \theta \| > cn^{-1/2}(\log n)^2\} = o(n^{-s/2}).$$

Let $F_n := X \times \Theta \to \mathbb{R}^k$ be $\mathscr{A} \times \mathscr{B}^k$-measurable. (The reader should not be confused by the fact that θ now stands for the x of Section 2.)

DEFINITION 2. A measurable map $T_n: X \to \mathbb{R}^k$ is an asymptotic F_n-estimator of order $(r, s) \in \mathbb{N}^2$ if it is an initial estimator of order s and if for every compact subset K of Θ there exist a constant $\gamma_K > 0$ such that

$$\sup_{\theta \in K} P_\theta\{x \in X : \|F_n(x, T_n(x))\| > (n^{-1/2}(\log n)^{\gamma_K})^{r+1}\} = o(n^{-s/2}).$$

We remark that these definitions generalize Pfanzagl's definitions (see [9, p. 212 and p. 249, respectively]).

General conditions for the existence of initial estimators may be obtained by the minimum distance method (see [8, p. 190, Lemma 2]). For practical purposes, however, it will be easier to use an initial estimator adjusted to the particular problem, e.g., the median as initial estimator for a location parameter and the range as initial estimator for the scale parameter.

We now give conditions which imply the existence of asymptotic F_n-estimators. First, we state the following general conditions:

Condition M_s. A function $H_n: X \times \Theta \to \mathbb{R}$ fulfills Condition M_s, if $x \to H_n(x, \theta)$ is $\mathscr{A}$-measurable for every $\theta \in \Theta$, and if there exist $h: \Theta \to \mathbb{R}$ such that for every compact subset K of Θ,

(i) $\sup_{\theta \in K} | h(\theta)| < \infty$

and

(ii) $\sup_{\theta \in K} P_\theta\{x \in X : | H_n(x, \theta) - h(\theta)| > 1\} = o(n^{-s/2})$.

We remark that in the i.i.d. case this condition is fulfilled for $H_n(x, \theta) = n^{-1} \sum_{i=1}^n H(x_i, \theta)$ if

$$\sup_{\theta \in K} E_\theta \mid H(\cdot, \theta)|^{s+\epsilon} < \infty.$$

Condition R_s. A function $H_n: X \times \Theta \to \mathbb{R}^k$ fulfills Condition R_s if for every compact subset K of Θ and every constant $c > 0$,

$$\sup_{\theta \in K} P_\theta\{x \in X : \| H_n(x, \theta)\| > cn^{-1/2} \log n\} = o(n^{-s/2}).$$

This condition is fulfilled in the i.i.d. case for $H_n(x, \theta) = n^{-1} \sum_{i=1}^n H(x_i, \theta)$ if $E_\theta H(\cdot, \theta) = 0$, the smallest eigenvalue of $\mathrm{Var}_\theta(H)$ is bounded away from zero on K, and $\sup_{\theta \in K} E_\theta \mid H(\cdot, \theta)|^{s+\epsilon} < \infty$.

Condition L_s. A matrix-function H_n on $X \times \Theta$ (i.e., $H_n: X \times \Theta \to (\mathbb{R}^k)^2$) fulfills Condition L_s if for every $\theta \in \Theta$ there exists a regular matrix $\Gamma(\theta)$ such that for every compact subset K of Θ and every constant $c > 0$,

$$\text{(i)} \quad \sup_{\theta \in K}(\| \Gamma(\theta)\| + \| \Gamma^{-1}(\theta)\|) < \infty$$

and

$$\text{(ii)} \quad \sup_{\theta \in K} P_\theta\{x \in X : \| H_n(x, \theta) - \Gamma(\theta)\| > cn^{-1/2} \log n\} = o(n^{-s/2}).$$

With these conditions and the results of Section 2 we immediately obtain our main result.

THEOREM. *Assume that* $\theta \to F_n(x, \theta)$ *is* $(r + 1)$*-times differentiable on* Θ *and that the function* $DF_n(x, \theta)$ *(differentiation with respect to* θ*) fulfills condition* L_s. *Assume furthermore that the partial derivatives* $D_{i_1 \dots i_j} F_n(x, \theta)$, $j = 3, \dots, r$, *fulfill Condition* M_s, *and that there exists an* $\mathscr{A}$*-measurable function* $h_n(\cdot, \theta): X \to \mathbb{R}$, $\theta \in \Theta$, *fulfilling Condition* M_s *such that for every* $\theta \in \Theta$ *there exists a neighborhood* U_θ *of* θ *with*

$$\sup_{\tau \in U_\theta} | D_{i_1 \cdots i_{r+1}} F_n(x, \tau)| < h_n(x, \theta), \; i_1, \dots, i_{r+1} \in \{1, \dots, k\}^{r+1}.$$

Assume, finally, that F_n *fulfills Condition* R_s. *Let* $T_n^{(0)}$ *be an initial estimator of order s. Then there exists*

(i) *an asymptotic* F_n*-estimator* T_n *of order* (r, s).

(ii) *a polynomial in* F_n *and its partial* θ*-derivatives, say* $S_n^{(r)}(x, \theta)$, *such that for every compact subset* K *of* Θ *and every asymptotic* F_n*-estimator* T_n *of order* (r, s) *there exist constants* $\epsilon_K, \gamma_K > 0$ *with*

$$\sup_{\theta \in K} P_\theta\{x \in X : \sup_{\|\tau - \theta\| < \epsilon_K n^{-1/2}(\log n)^2} \| T_n(x) - S_n^{(r)}(x, \tau)\|$$
$$> (n^{-1/2}(\log n)^{\gamma_K})^{r+1}\} = o(n^{-s/2}).$$

Proof. The proof immediately follows from the remarks at the beginning of Section 2, Lemmas 2, 4, and Remark 3 in Section 2.

(We only need to check that there exists a set $A_{n,\theta} \in \mathscr{A}$ with

$$\sup_{\theta \in K} P_\theta(A_{n,\theta}^c) = o(n^{-s/2})$$

such that for $x \in A_{n,\theta}$ and n sufficiently large all conditions of the lemmas are fulfilled if $\epsilon_K > 0$ is sufficiently small, and $\gamma_K > 0$ is sufficiently large. Obviously the true parameter θ stands for x_0 of Section 2.)

REFERENCES

[1] CHIBISOV, D. M. (1972). An asymptotic expansion for the distribution of a statistic admitting an asymptotic expansion. *Theor. Prob. Appl.* **17** 620–630.

[2] CHIBISOV, D. M. (1973). An asymptotic expansion for a class of estimators containing maximum likelihood estimators. *Theor. Prob. Appl.* 18 295–303.

[3] FISHER, R. A. (1925). *Statistical Methods for Research Workers.* Oliver and Boyd, Edinburgh.

[4] LECAM, L. (1956). On the asymptotic theory of estimation and testing hypotheses. *Proc. Third Berkeley Symp. Math. Statist. Prob.*, Vol. 1, pp. 129–156.

[5] LINNIK, Y. V., AND MITROFANOVA, N. M. (1965). Some asymptotic expansions for the distribution of the maximum likelihood estimate. *Sankhyā* **27A** 73–82.

[6] MICHEL, R. (1975). An asymptotic expansion for the distribution of asymptotic maximum likelihood estimators of vector parameters. *J. Multivar. Anal.* **5** 67–82.

[7] MITROFANOVA, N. M. (1967). An asymptotic expansion for the maximum likelihood estimate of a vector parameter. *Theor. Prob. Appl.* **12** 364–372.

[8] PFANZAGL, J. (1972). Further results on asymptotic normality, *I. Metrika* **18** 174–198.

[9] PFANZAGL, J. (1973). Asymptotically optimum estimation and test procedures. *Proceedings of the Prague Symposium on Asymptotic Statistics* (J. Hájek, Ed.), Vol. 1, pp. 201–272.